

Abstract of the Disclosure

The invention provides a method and system for caching information objects transmitted using a computer network. A cache engine determines directly when and where to store those objects in a memory (such as RAM) and mass storage (such as one or more disk drives), so as to optimally write those objects to mass storage and later read them from mass storage, without having to maintain them persistently.

The cache engine actively allocates those objects to memory or to disk, determines where on disk to store those objects, retrieves those objects in response to their network identifiers (such as their URLs), and determines which objects to remove from the cache so as to maintain sufficient operating space. The cache engine collects information to be written to disk in write episodes, so as to maximize efficiency when writing information to disk and so as to maximize efficiency when later reading that information from disk. The cache engine performs write episodes so as to atomically commit changes to disk during each write episode, so the cache engine does not fail in response to loss of power or storage, or other intermediate failure of portions of the cache. The cache engine also stores key system objects on each one of a plurality of disks, so as to maintain the cache holographic in the sense that loss of any subset of the disks merely decreases the amount of available cache. The cache engine also collects information to be deleted from disk in delete episodes, so as to maximize efficiency when deleting information from disk and so as to maximize efficiency when later writing to those areas having former deleted information. The cache engine responds to the addition or deletion of disks as the expansion or contraction of the amount of available cache.